Part 3

Why Gene Duplication?

Chapter X

Duplication for the Sake of Producing More of the Same

The discussions presented in Part 2 revealed the true character of natural selection. It is not so much an advocator or mediator of heritable changes, but rather it is an extremely efficient policeman which conserves the vital base sequence of each gene contained in the genome. As long as one vital function is assigned to a single gene locus within the genome, natural selection effectively forbids the perpetuation of mutations affecting the *active* sites of a molecule. In the case of the enzyme locus, *tolerable* mutations might change the kinetic property such as pH optimum and Michaelis constant of the enzyme, but never the basic character. Therefore, the dihydro-orotase locus would forever remain the dihydro-orotase locus, and the β-galactosidase locus would remain the β-galactosidase locus.

It becomes quite clear that while allelic changes at already existing gene loci suffice for racial differentiation within species as well as for adaptive radiation from an immediate ancestor, they cannot account for large changes in evolution, because large changes are made possible by the acquisition of new gene loci with previously non-existent functions. Only by the accumulation of *forbidden* mutations at the *active* sites can the gene locus change its basic character and become a new gene locus. An escape from the ruthless pressure of natural selection is provided by the mechanism of gene duplication. By duplication, a redundant copy of a locus is created. Natural selection often ignores such a redundant copy, and, while being ignored, it accumulates formerly *forbidden* mutations and is reborn as a new gene locus with a hitherto non-existent function. Thus, gene duplication emerges as the major force of evolution.

Even before the advent of molecular biology, a number of geneticists with foresight, such as HALDANE (1932), realized the important role gene duplication played in evolution. However, full appreciation of the magnitude of importance was not possible until the elucidation of the coding mechanism enabled us to interpret evolutional changes reflected in the direct gene products.

Although the creation of new gene loci by supplying redundancy is the most important role, there are other benefits the mechanism of gene duplication confers to organisms. When the metabolic requirement of an organism dictates the presence of

an enormous amount of a particular gene product, the incorporation of multiple copies of the gene locus by the genome often fulfills that requirement. This, then, is the type of gene duplication which serves to produce more of the same gene product.

1. Genes for *Ribosomal* RNA

As stated in Chapter II, an organism only requires three different kinds of *ribosomal* RNA; 5S, 18S and 28S. While *ribosomal* RNA is short in variety, it has to be made in great quantity, for the translation of a single *messenger* RNA requires the attachment of several ribosomes, and a single cell is likely to be producing many copies each of hundreds of different kinds of *messenger* RNA. Thus, as much as 85% of the total RNA extracted from ordinary somatic cells is *ribosomal* RNA.

Quite clearly, if the genome contains only a single DNA cistron for each of the three different kinds of *ribosomal* RNA, individual somatic cells cannot synthesize enough *ribosomal* RNA to sustain the ontogenic development. Using the technique of DNA-RNA hybridization already mentioned in Chapter II, RITOSSA and SPIEGELMAN (1965) have shown in the fruit fly *(Drosophila melanogaster)* that each nucleolar organizer of this insect species contains 100 tandemly duplicated copies of a pair of genes which transcribe for a dicistronic RNA which is later split into 18S and 28S *ribosomal* RNA. In the fruit fly, the nucleolar organizer is carried by the X as well as the Y-chromosome.

In the case of the African water frog *(Xenopus laevis)*, the nucleolar organizer is carried by a single pair of homologous autosomes. The latest estimate indicates that in this vertebrate species each nucleolar organizer contains 450 tandemly duplicated copies of a pair of genes for 18S and 28S *ribosomal* RNA (BROWN and DAWID, 1968). However, it should be remembered that the genome size (the haploid DNA content) of *Xenopus laevis* is 30 or 40 times greater than the genome size of *Drosophila melanogaster*; therefore, it appears that in proportion to the genome size, *Drosophila* has a greater number of genes for two classes of *ribosomal* RNA. The genome size of *Xenopus* is only slightly smaller than that of mammals. But, in the case of mammalian species, several pairs rather than a single pair of chromosomes tend to carry nucleolar organizers. For instance, of the 46 chromosomes of man, five different pairs of acrocentric autosomes carry nucleolar organizers [Fig. 3 (Plate I), Chapter III]. Could it be that evolution from cold-blooded to warm-blooded vertebrates was accompanied by an increase in the degree of duplication of a *ribosomal* RNA cistron? It is likely that a higher rate of metabolism requires greater concentration of ribosomes in the cell.

It has been shown that a gene for the third class of *ribosomal* RNA (5S) is not contained in the nucleolar organizing region of the chromosome either in *Drosophila* or in *Xenopus*. However, there appears to be extreme redundancy of 5S DNA in the genome. The latest estimate is that the genome of *Xenopus* contains 20,000 duplicated copies of a DNA cistron for 5S *ribosomal* RNA (BROWN and DAWID, 1968).

Aside from the three classes of RNA mentioned above, ribosomes also contain proteins. There is little doubt that for the continuous formation of ribosomes, the cell has to synthesize as much *ribosomal* protein as *ribosomal* RNA. It is of extreme interest to find out whether or not natural selection also favored the amplification by tandem duplication of each structural gene for *ribosomal* protein. In the case of *Escherichia coli*, at least 16 different kinds of *ribosomal* proteins, ranging in molecular weight from 9,000 to 41,000, have been identified (KALTSCHMIDT *et al.*, 1967).

Even further amplification of the genes for 18S and 28S *ribosomal* RNA appears to occur during oögenesis of amphibians and echinoderms. As mentioned earlier, an individual *Xenopus*, which is homozygous for a deletion of the nucleolar organizer is totally incapable of synthesizing 18S and 28S *ribosomal* RNA. Yet such a homozygous deficient zygote derived from the mating of heterozygotes grows to the swimming tadpole stage (ELSDALE *et al.*, 1958). The amount of *ribosomal* RNA stored in the egg cytoplasm by a heterozygous mother is sufficient to sustain the growth of homozygous embryos to this advanced stage of development. It is clear that even the nucleolar organizer, with 450 copies of a *ribosomal* RNA gene, cannot, by itself, produce such enormous amounts of 18S and 28S RNA during oögenesis. It is now shown that as the oöcyte suspended in the diplotene stage of first meiotic prophase begins to grow in size, the nucleolar organizer region of the chromosome disseminates its free copies to the nuclear plasm, so that the oöcyte nucleus finally contains 1000 or so free copies of the nucleolar organizers; each of which appears to organize an individual nucleolus. Since each nucleolar organizer already contains 450 tandemly duplicated copies of a pair of genes for 18S as well as 28S *ribosomal* RNA, the number of genes for two classes of *ribosomal* RNA which become available to the growing oöcyte is truly staggering; 450×1002. In sharp contrast to 18S and 28S, free copies of a gene for 5S *ribosomal* RNA do not appear to be disseminated during oögenesis of *Xenopus*. As the chromosomes already contain 20,000 duplicates of this gene, further amplification appears unnecessary (BROWN and DAWID, 1968).

In the case of amniote eggs of reptiles, birds, and mammals, such dissemination of free copies of the nucleolar organizer during oögenesis probably occurs on a much smaller scale, if it occurs at all. Nevertheless, the fact that a segment of the chromosome can engage in repeated DNA replication and disseminate its free copies, while the rest of the chromosomes are not involved in DNA replication, has far reaching implications.

2. Genes for *Transfer* RNA

The *transfer* RNA is also short in variety, for the genome of an organism needs to contain only 30 or so different kinds of genes which transcribe *transfer* RNA (Chapter II). However, individual cells need to produce each species of *transfer* RNA in rather large amounts, for a single translation of a *messenger* RNA of average length to a polypeptide chain requires a few hundred *transfer* RNA. Indeed, as much as 15% of the total RNA extracted from the growing embryo is *transfer* RNA. One might expect that in the case of *transfer* RNA too, natural selection favored the tandem duplication of a cistron for each species of *transfer* RNA. Again utilizing the technique of DNA-RNA hybridization, RITOSSA, ATWOOD and SPIEGELMAN [1966 (1)] arrived at the conclusion that if there are 60 different kinds of genes for *transfer* RNA, the genome of *Drosophila melanogaster* contains 13 duplicated copies of each gene.

In the case of *transfer* RNA genes, however, there remains some doubt as to whether or not the apparent redundancy revealed by the DNA-RNA hybridization can really be taken as evidence of the genome containing multiple replicas of each *transfer* RNA gene. The genome of *E. coli* contains two rather widely separated gene loci (Su2 and Su3) or gene clusters for two subspecies of tyrosine *transfer* RNA. One subspecies specified by Su3 has the anticodon which can recognize both codons (UAU, UAC) for tyrosine. Thus, it would appear that the other subspecies specified

by Su2 must also have the same anticodon. Yet these two tyrosine *transfer* RNA's are not identical to each other (GAREN, 1968). Su2 and Su3 of *E. coli* should be regarded as two closely related but separate gene loci diverged from a common ancestral gene after duplication, rather than exact replicas of each other. However, the fact that a subspecies specified by Su3 amounts to only 10% of the total tyrosine *transfer* RNA can be explained on the basis that while Su2 represents a cluster of 10 tandem duplicates, Su3 is a singleton.

3. Inherent Disadvantage of Having Multiple Copies of the Same Gene

On the surface, it would appear that whenever the need arises for an organism to have an enormous amount of one particular gene product, this need can easily be satisfied by incorporating the multiple copies of the same gene into the genome. In fact, the nature of natural selection and chromosomes are such that the incorporation of the multiple copies entails inherent disadvantages.

The fact that *ribosomal* RNA isolated from *Xenopus laevis*, of the tailless amphibian order *Anura*, hybridize very well with the nucleolar organizing DNA isolated from salamanders, such as *Axolotl mexicanum* and *Necturus maculosus*, reveals that natural selection has stringently conserved the base sequence of a pair of genes for 18S as well as 28S *ribosomal* RNA (BROWN and DAWID, 1968). Anurans and salamanders followed separate paths of evolution for as long as 280 million years (since the first amphibians emerged on this earth at the beginning of the Carboniferous period). If base substitutions are permissible at many of the sites of this cistron, nearly 300 million years of separation would have resulted in a marked difference in the base sequences between the anuran nucleolar organizing DNA and the salamander nucleolar organizing DNA. Thus, anuran *ribosomal* RNA would not effectively hybridize with the salamander nucleolar organizing DNA.

Natural selection can eliminate *forbidden* mutations and effectively police the base sequence of a DNA cistron only if the genome contains a single copy of each gene. Policing by natural selection becomes very ineffective when multiple copies of the gene are present. For the sake of simplicity, let us assume that the genome contains three exact replicas of the same gene. A *forbidden* mutation which rendered one of the three copies functionless would be tolerated, since even a deficient homozygote still has four doses of the good gene. The second *forbidden* mutation which renders the second copy useless would also be tolerated, for even a deficient double homozygote still has two doses of a good gene. Thus, in a relatively short time, two of the three duplicates would join the ranks of "garbage DNA", and finally only one functional gene remains in the genome. Consequently, having hundreds of tandemly duplicated copies of a single gene in the nucleolar organizer is not as ideal a situation as it appears on the surface, for slowly but surely more and more duplicates would become useless genes by mutation. Ideally, the gamete should contain only a single gene each for 18S and 28S *ribosomal* RNA, and the tandem duplication of it should occur only after fertilization. This way, all multiple copies of a *ribosomal* gene contained in an individual are either uniformly defective or uniformly functional. Natural selection can now eliminate unfit individuals which inherited a defective *ribosomal* gene. It is surprising that no known organism employs this ideal solution for the policing of its *ribosomal* genes.

CALLAN (1967) has proposed a very ingenious mechanism by which the organism might escape the hazard of containing multiple copies of the gene in the genome He postulates that there is a hierarchy among the tandem duplicates in that the one at the end is the master, while all others are slaves; the master-slave theory. When chromosomes duplicate before each cell division, not the slaves, but only the master serves as the template for DNA replication. The net effect of the master-slave system is the same as the gamete having only a single dose of the *ribosomal* gene, since all the *ribosomal* genes contained in an individual are either uniformly defective or uniformly

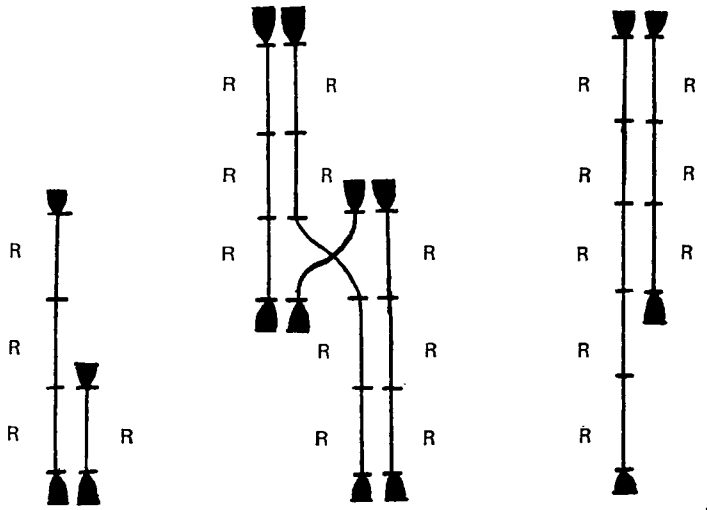

Fig. 9. The consequence of unequal crossing-over between duplicated segments is illustrated on the nucleolar organizer which normally contains three tandemly duplicated copies of a pair of genes (R) for 18S and 28S *ribosomal* RNA. Middle column: First meiotic prophase. The homologous pairing between the duplicated segments is inexact. As a result, a chiasma is exchanged between the third gene of the chromosome at the left and the first gene of its homologue at the right. Left and right columns: Two daughter cells in 2nd meiosis. At the left, one of the two chromatids received the deleted nucleolar organizer (one R). At the right, one of the two received the further duplicated nucleolar organizer (five R's). As each crossing-over involves two of the four chromatids of the two homologues in pairing, two of the four gametes produced are affected.

functional. If the master suffers a *forbidden* mutation, all the slaves of the next cell generation would inherit the same defect.

One wonders if it is this master-slave system which enabled anurans and salamanders to stringently conserve the base sequence of the *ribosomal* genes despite the presence in their genome of 450 or so tandemly arranged copies.

Another serious difficulty an organism encounters by having multiple copies of the same gene is constant deletion and further duplication which afflicts the chromosomal region made of tandemly duplicated copies. Crossing-overs that normally occur between homologous chromosomes during meiosis are, as a rule, no problem, for exchanges are preceded by exact gene-for-gene pairing between two homologous chromosomes. In the case of a duplicated region, however, homologous pairing becomes very inexact. For instance, No. 1 *ribosomal* gene at the head of the nucleolar organizing region of one chromosome might pair with No. 250 *ribosomal* gene in the

middle of the nucleolar organizer of its homologue. The result of such shifted pairing is "unequal crossing-over". Where both chromosomes had the nucleolar organizer made of 450 copies of a *ribosomal* gene, one would now receive only 200 copies (deletion), while the other receives 700 copies (further duplication) as shown in Fig. 9. If homologous pairing is truly based on the attraction that exists between the DNA of nearly identical base sequences, such shifted pairing and subsequent unequal exchange should also occur between two chromatids of the same chromosome in somatic cells. Such unequal crossing-over and unequal exchange between the nucleolar organizers on the X and the Y are constantly occurring in the fruit fly *(Drosophila melanogaster)*. Those which received the considerably deleted nucleolar organizers from both parents finally become recognizable because of their markedly retarded growth [RITOSSA *et al.*, 1966 (2)]. Those affected flies have been known as *bobbed* mutants (STERN, 1927). Because further unequal crossing-over between the deleted nucleolar organizers occasionally result in restoration of the normal nucleolar organizer, normal flies frequently emerge from a stock of *bobbed* mutants. The reciprocal product of unequal crossing-over is the extraordinarily large nucleolar organizer containing a greater than normal number of duplicates of a *ribosomal* gene. Contrary to what one might expect, a fly which inherited such a great nucleolar organizer does not become a superfly.

There is yet another class of mutations in *Drosophila* which result in generalized growth retardation. They are known as *Minutes*, for they are homologous lethal, dominant traits. Although the *Minutes* form a phenotypically homogeneous class, any of the over 50 independent gene loci widely scattered in the genome can mutate to become a *Minute*. ATWOOD [in RITOSSA *et al.*, 1966 (1)] postulates that each *Minute* is also a deletion due to unequal crossing over affecting one cluster of 13 duplicates of a particular *transfer* RNA gene.

Such deleterious consequence of unequal crossing-over is the fate which has to be endured by the chromosomal segment carrying the tandem duplicates of the same gene. Yet, in the absence of either of the two ideal systems, one where the gamete contains only one dose of the gene with duplication occurring after fertilization and the other, the master-slave system, apparently deleterious deletions might be beneficial to the species in the long-run. As a result of deletion, the nucleolar organizer can cleanse itself of degenerate duplicates which became functionless due to accumulation of mutations. Subsequent unequal crossing-over between the partially deleted nucleolar organizers can restore the original degree of duplication this time made mostly of functional copies.

For mammalian species which carry the nucleolar organizers on several different chromosomes, the additional problem of maintaining homology between the regions of non-homologous chromosomes is imposed. Unless all these nucleolar organizing regions involve themselves in mutual exchange of genetic materials, some would become a useless collection of degenerate copies not contributing to the production of 18S and 28S *ribosomal* RNA.

Of 46 chromosomes in the diploid nucleus of man, the nucleolar organizers are carried by the three pairs of medium-sized acrocentric autosomes (13th, 14th and 15th pairs) as well as by the two smallest pairs of acrocentric autosomes (21st and 22nd pairs) [Fig. 3 (Plate 1), Chapter III]. In human somatic cells, all these acrocentric autosomes are often in very close association with each other at their nucleolar organ-

izers (FERGUSON-SMITH and HANDMAKER, 1961; OHNO *et al.*, 1961). This appears to be a mean employed by mammals to maintain the homology between the nucleolar organizers carried by non-homologous chromosomes.

References

BROWN, D. D., DAWID, I. B.: Specific gene amplification in oöcytes. Science **160**, 272—280 (1968).

CALLAN, H. G.: The organization of genetic units in chromosomes. J. Cell Sci. **2**, 1—7 (1967).

ELSDALE, T. R., FISCHBERG, M., SMITH, S.: A mutation that reduces nucleolar number in *Xenopus laevis*. Exptl. Cell Res. **14**, 642—643 (1958).

FERGUSON-SMITH, M. A., HANDMAKER, S. D.: Observations on the satellited human chromosomes. Lancet **1961 I**, 638—640.

GAREN, A.: Sense and nonsense in the genetic code. Science **160**, 149—159 (1968).

HALDANE, J. B. S.: The causes of evolution. New York: Harper and Bros. 1932.

KALTSCHMIDT, I., DZIONARA, M., DONNER, D., WITTMANN, H. G.: Ribosomal proteins. Mol. Gener. Genet. **100**, 364—373 (1967).

OHNO, S., TRUJILLO, J. M., KAPLAN, W. D., KINOSITA, R.: Nucleolus-organizers in the causation of chromosomal anomalies in man. Lancet **1961 II**, 123—125.

RITOSSA, F. M., SPIEGELMAN, S.: Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. US **53**, 737—745 (1965).

— ATWOOD, K. C., SPIEGELMAN, S.: (1) On the redundancy of DNA complementary to amino acid transfer RNA and its absence from the nucleolar organizing region of *Drosophila melanogaster*. Genetics **54**, 663—676 (1966).

— — — (2) A molecular explanation of the *bobbed* mutants of *Drosophila* as partial deficiencies of "ribosomal" DNA. Genetics **54**, 819—834 (1966).

STERN, C.: Ein genetischer und zytologischer Beweis für Vererbung in Y-chromosome von *Drosophila melanogaster*. Z. Induktive Abstammungs- u. Vererbungslehre **44**, 187—231 (1927).

Chapter XI

The Attainment of a Permanent Heterozygous Advantage by the Incorporation of Two Former Alleles into the Genome

The heterozygous advantage, which benefits only certain members of a population, can be fixed as a new species characteristic if two alleles involved are incorporated into the genome as two separate gene loci. In this way, every member of a species would come to enjoy the heterozygous advantage without ever having to produce homozygotes which may be unfit.

Let us imagine a hypothetical species of fish which inhabits a long stretch of a river. This river originates in a high Northern mountain and runs through a Southern desert before it pours out to the ocean. Further, let us assume that this species is endowed with two allelic alternatives at the gene locus for an enzyme; esterase (Es). The Es A-variant specified by one allele has the temperature optimum of 5 °C, while the Es B-variant specified by the other allele functions best at 20 °C. It is expected that for a subpopulation inhabiting the high Northern mountain part, natural selection unconditionally favored the A-variant, so that the subpopulation as a whole

has become A/A homozygous. Conversely, the B-variant has been favored in the other subpopulation inhabiting the low Southern desert part, and the B should have become the wild-type allele of that subpopulation. The water temperature in the intermediate part of the river fluctuates widely with seasons; very cold in winter and quite hot in summer. There is an unquestionable heterozygous advantage for members of a subpopulation occupying the intermediate area. The A/B heterozygote can cope with both the cold temperature of winter and the hot temperature of summer. Yet as long as the A- and B-variants are specified by two alleles of the same gene locus, only 50% of the subpopulation in the intermediate area can become heterozygotes. Twenty-five per cent of the zygotes would be A/A which have difficult summers, and the other 25%, which are B/B, would encounter serious problems of survival during the winter months. Under these conditions, one would expect that natural selection has favored the duplication of the Es-locus in the intermediate subpopulation. When two alleles for the A- and B-variants are incorporated into the genome as two separate gene loci, every member of the subpopulation would become AB/AB and enjoy a permanent heterozygous advantage without ever having to produce undesirable homozygotes. A close approximation of this hypothetical situation has apparently occurred in populations of the catostomid fish *(Catostomus clarki)* which inhabit tributaries of the Colorado River system (KOEHN and RASMUSSEN, 1967).

Whenever natural selection strongly selects against homozygotes, a duplication which confers the heterozygous advantage to every member of a population must be favored. Incorporation of two former alleles into the genome, however, contains the germ of disaster. No gene functions alone; rather a group of genes perform interrelated functions. For example, glucose-6-phosphate dehydrogenase (G-6-PD) and 6-phosphogluconate dehydrogenase (6-PGD) catalyze two successive steps of the pentose phosphate shunt of carbohydrate metabolism. Once the activities of these two enzymes of the species have been coordinated on the basis of a one-to-one gene dosage ratio (two-to-two in diploid somatic cells), duplication of the 6-PGD locus without concordant duplication of the G-6-PD locus might be disastrous. Thus, even if there is strong natural selection against homozygotes at the 6-PGD locus, incorporation of two 6-PGD alleles into the genome might not be permitted, for the disadvantage of disrupting the established gene dosage relationship with G-6-PD might outweigh the advantage to be gained by the duplication. Hemoglobin α- and β-chains are specified by two unlinked genes. Yet two alpha's and two beta's together make up a single hemoglobin molecule. The fact that a duplication which incorporates the normal β-chain and mutant β'-chain genes into the genome has not occurred in African populations despite a strong heterozygous advantage might indicate that having two doses of the β-chain gene while maintaining only a single dose of the α-chain gene is incompatible with proper ontogenic development. In diploid organisms, the gene dosage appears to be of prime importance. Otherwise, mammals would not have developed the elaborate dosage compensation mechanism for X-linked genes as discussed in Chapter III.

It becomes clear that despite the obvious benefit of attaining a permanent heterozygous advantage, this type of duplication cannot always be favored. The concordant duplication of all genes with interrelated functions which are scattered over different chromosomes in the genome is accomplished only by becoming tetraploid. This is

the very reason we believe that polyploidy played just as important a role in the evolution of vertebrates as it did in the evolution of higher plants. This point shall be discussed in detail in Part 4.

References

KOEHN, R. K., RASMUSSEN, D. I.: Polymorphic and monomorphic serum esterase heterogeniety in Catostomid fish populations. Biochem. Genet. 1, 131—144 (1967).

Chapter XII

The Differential Regulation of Former Alleles and Their Transformation to Isozyme Genes

Nonconcordant duplication involving only one of a group of functionally interrelated gene loci becomes permissible if the incorporation of two former alleles of that locus into the genome is quickly followed by the development of the differential genetic regulatory mechanism. As this genetic regulatory mechanism permits only one or the other former allele to engage in transcriptional activity in any given somatic cell type of an individual, the original one-to-one gene dosage relationship is effectively restored among all functionally related genes in spite of discordant duplication which had affected one locus.

Once it is possible for an organism to discriminate between duplicated genes for the same enzyme and use them differentially during ontogenic development, the way is open for an organism to derive ultimate benefit from this type of gene duplication. Because of differential use, the duplicated genes are exposed to different pressures of natural selection. As a result, the two would gradually diverge from each other by accumulating different kinds of mutations. Finally, the products of the duplicated genes, although they still act upon the same substrate and use the same coenzyme, acquire kinetic properties which are markedly different from each other. In such a way, a group of duplicated genes for the so-called isozymes must have been born.

Most vertebrates are endowed with at least two separate gene loci for A and B-subunits of an enzyme, lactate dehydrogenase (LDH). These different subunits can recognize each other as well as themselves. Accordingly, by polymerization, five tetrameric isozymes are formed between two kinds of subunits. They are A_4, A_3B, A_2B_2, AB_3 and B_4 (MARKERT, 1964). The fact that the products of two separate gene loci maintain infinite affinity for each other suggests that the two arose from a common ancestral gene by duplication. While subsequent natural selection could easily have maintained the already existing affinity between the two subunits, it would have been much more difficult to create an affinity between the products of two separate gene loci which had no initial affinity.

Similarly, mammals and other vertebrates are endowed with three separate gene loci for A, B and C-subunits of an enzyme; fructose diphosphate aldolase. In the case of this enzyme, two subunits are seldom produced in the same tissue. For instance, muscle cells produce only A-subunits while liver cells produce only B-subunits.

However, A- and B-subunits mixed *in vitro* randomly polymerize with each other and form five tetrameric isozymes in the same manner as do A- and B-subunits of LDH (PENHOET *et al.*, 1966). Since the high affinity between A and B subunits is seldom utilized by an organism, in the case of aldolase, this affinity probably reflects the fact that the three separate gene loci for aldolase are duplicates derived from a single gene locus.

What is the nature of the differences in kinetic properties of the products of these duplicated gene loci and how are these differences exploited by an organism during ontogenic development? In the case of LDH which catalyzes the interconversion of lactate and pyruvate accompanied by the interconversion of $NADH_2$ and NAD, it has been shown that LDH-5 which is made exclusively of A-subunits (A_4) has a low affinity for pyruvate. That is to say, it functions most efficiently when the substrate (pyruvate) concentration is around 10^{-3} M. In sharp contrast, LDH-1 which is made exclusively of B-subunits (B_4) has a high affinity; it functions best when the pyruvate concentration is around 2×10^{-4} Mol (PLAGEMANN *et al.*, 1960). Needless to say, LDH-2, 3 and 4, which are hybrid molecules made of A- and B-subunits, have intermediate kinetic properties. While pyruvate occupies the key position in carbohydrate metabolism, lactate appears to serve no useful purpose in high organisms other than as a temporary electron acceptor or oxidant during periods when oxygen is in short supply. Thus, the most important function of LDH may not be the reduction of pyruvate or the oxidation of lactate, but rather the regulation of the ratio of NAD to $NADH_2$, since this ratio affects the rates of many catalytic reactions (MARKERT, 1964).

From the above, it is easy to see that LDH-5 (A_4) is most useful to the tissues which are anaerobic because of a relatively poor blood supply. During the rapid metabolism of glucose, pyruvate production is enhanced and NAD is rapidly reduced to $NADH_2$. In the absence of oxygen, $NADH_2$ cannot be reoxidized to NAD, and, unless something is done, glycolysis soon comes to a grinding halt. The presence of low affinity LDH-5 (A_4) enables $NADH_2$ to be reoxidized to NAD by the conversion of pyruvate to lactate. This metabolic arrangement permits continued energy production under the anaerobic condition even to the point of toxic accumulation of lactic acid. On the other hand, in the case of the tissues which are well oxygenated by an abundant blood supply, LDH-1 (B_4), which has a high affinity for pyruvate, is no doubt the preferred type. For the tissues which are periodically exposed to the anaerobic condition, the concommitant presence of both A- and B-subunits would be preferable, for the majority of LDH molecules produced would then be the hybrid types; LDH-2, 3 and 4.

In the case of mammals, LDH-5 (A_4) indeed predominates in all fetal tissues, for mammalian fetuses lead a rather anaerobic existence. After birth, LDH-5 remains predominant in skeletal muscle which has a poor blood supply and where lactic acid accumulates to an alarming degree after strenuous exercise. In the well oxygenated tissues, notably the heart, the production of LDH-5 is suppressed after birth and the active transcription and translation of the gene for B-subunits begins, so that LDH-1 predominates in the postnatal heart (MARKERT, 1964).

In the case of aldolase too, the kinetic property of the A_4-molecule must accommodate the unique metabolic requirement of skeletal muscle, while that of the B_4-molecule fulfills the different metabolic requirement of hepatic cells (PENHOET *et al.*,

1966). It is becoming increasingly clear that in higher organisms such as vertebrates the task of specifying one particular enzyme type is more often assigned to a group of duplicated gene loci rather than to a single gene locus. Thus, man and other mammals are endowed with at least two separate gene loci for pyruvate kinase (PK) (KOLER *et al.*, 1964) and with three unlinked gene loci for phosphoglucomutase (PGM) (HARRIS *et al.*, 1967).

Starting from a single fertilized egg, the body of vertebrates becomes a complex organization made of hundreds of different kinds of somatic cell types, and no two somatic cell types are identical with regard to their assigned functions. There is little doubt that the type of gene duplication discussed in this chapter contributed greatly to the attainment of such a complex body organization. Cells having identical genetic material can differentiate into different somatic cell types only because the genome contains a group of duplicated gene loci for each of many key enzymes. Although duplicated genes of the group specify the same enzyme so far as the choice of substrate and coenzyme is concerned, each gene product is unique with regard to its Km as well as its pH and temperature optimum. Because the choice from each group of duplicates is offered, different somatic cell types acquire different characteristics even with regard to the process of basic carbohydrate metabolism.

The production of specialized non-enzymatic proteins is also benefited from this type of gene duplication. Each immunoglobulin molecule is made of light-chains (L-chains) and heavy-chains (H-chains), and L- and H-chains are specified by unlinked groups of duplicated genes. Man and other mammals apparently carry only two separate gene loci for κ- and λ-classes of L-chains on one chromosome, and as many as 10 separate gene loci for various classes of H-chains on another chromosome. Natural selection has permitted such gross discordance with regard to the degree of duplication of L- and H-chain genes only because the genetic regulatory mechanism which insures that only a single L-chain locus and a single H-chain locus shall engage in transcription in each clone of antibody-producing plasma cells had previously been evolved (PUTNAM *et al.*, 1967; NATVIG *et al.*, 1967; HERZENBERG *et al.*, 1967; POTTER and LIEBERMAN, 1967; COHN, 1967; HOOD *et al.*, 1968). Having a group of functionally diverged genes for H-chains enables mammals to cope with all sorts of contingencies. The IgA-type of antibody made up of an α-class H-chain is secreted into the milk by the mother, and the maternally derived IgA protects newborns while they themselves are still incapable of antibody production. IgM-type of 19S antibodies, which are polymeric molecules, use the μ-class of H-chains. IgM fixes complements and contributes greatly to the body's initial reponse to newly encountered antigens. IgG-type of antibodies use the γ-class of H-chains, and this is the cornerstone of the body's defense against invading foreign organisms. Furthermore, IgG of the mother passes through the placenta and protects the embryo.

Although there are numerous examples of functionally diverged duplicated genes in vertebrates, most duplications appear to have occurred in ancient times, so that even rather remotely related species usually show the same degree of gene duplication and the same differential use of these duplicated genes. It is not often that one is afforded with an opportunity to catch the type of gene duplication described in this and the previous chapter in the middle of the act so to speak. Fortunately, duplication of the 6-PGD locus which occurred in certain members of the fish family *Cyprinidae* afforded us with such an opportunity.

Minnows, carp and goldfish comprise this fresh water family. While most of the minnows are diploid species having the mean 2n number of 52, *Barbus barbus* of the Rhine River in Europe as well as the familiar goldfish and carp appear to be tetraploid species for they have more than 100 chromosomes. The genome of many diploid species contains a single autosomally inherited gene locus for this enzyme, and each species maintains a number of alleles at this locus which specify electrophoretic variants of 6-PGD. Since each 6-PGD molecule is a dimer, the homozygote such as A/A of the diploid species gives a single band of this enzyme (A_2) upon starch gel electrophoresis, while the heterozygote such as A/B shows three enzyme bands (A_2, AB and B_2) in the expected 1:2:1 ratio (Fig. 10). As a result of duplication of the entire genome, the tetraploid species have apparently incorporated two former alleles of the 6-PGD

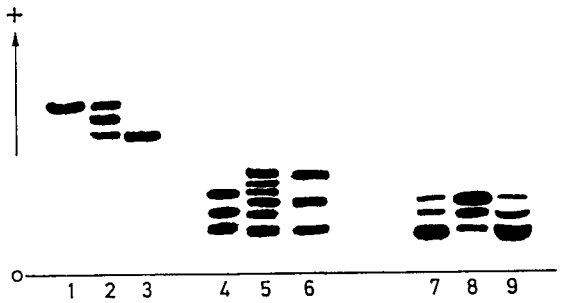

Fig. 10. The starch gel plate stained for 6-PGD. Vertical starch gel electrophoresis was carried out at pH 8.6 using a borate buffer. The starting point is indicated by zero, and the anodal direction is upward. *1, 2 and 3* illustrate the allelic polymorphism of a single gene locus in a diploid species, *Barbus tetrazona*. (1) A single C_2 band of a C/C homozygote. (2) Three bands of an A/C heterozygote. A hybrid dimer band in the middle is intensely stained. (3) A single A_2 band of an A/A homozygote. *4, 5, and 6* illustrate the allelic polymorphism of one of the two gene loci in the goldfish which is a tetraploid species. (4) Three bands of an A/A, B^2/B^2 homozygote. (5) Six bands of an A/A, B^3/B^2 homozygote. (6) Three bands of an A/A, B^3/B^3 homozygote. *7, 8 and 9* illustrate differences in the isozyme patterns of different tissues of the goldfish which is homozygous A/A, B^2/B^2. (7) Gills where A_2 band at the bottom is most conspicuous. (8) Liver where B_2^2 band at the top is accentuated. (9) Kidney where A_2 band is again most conspicuous

locus as two separate gene loci. Accordingly, the double homozygote such as A/A, B/B of the tetraploid species shows the three enzyme band pattern reminiscent of the diploid heterozygote. The tetraploid individual which is heterozygous at one of the two gene loci such as A/A, B^3/B^2 now shows six 6-PGD bands; A_2, AB^2, AB^3, B_2^2, B^2B^3 and B_2^3 (Fig. 10). Thus, the tetraploid cyprinid fish have indeed attained a permanent heterozygous advantage by the incorporation of two former alleles into the genome as two separate gene loci. Furthermore, when different tissues of the tetraploid species are compared, it is noted that in certain tissues such as gills and kidney there is a predominance of the A-subunits, while in other tissues such as the liver, more B-subunits than A-subunits were apparently produced (Fig. 10). It appears that these tetraploid species are on the way to developing the differential genetic regulatory mechanism which discriminates between the two former alleles. The gene loci for A- and B-subunits of 6-PGD are indeed becoming the isozyme genes as described in this chapter (Bender and Ohno, 1968).

References

BENDER, K., OHNO, S.: Duplication of the autosomally inherited 6-phosphogluconate dehydrogenase gene locus in tetraploid species of Cyprinid fish. Biochem. Genet. **2**, 101—107 (1968).

COHN, M.: Natural history of the myeloma. Cold Spring Harbor Symposia Quant. Biol. **32**, 211—222 (1967).

HARRIS, H., HOPKINSON, D. A., LUFFMAN, J. E., RAPLEY, S.: Electrophoretic variation in erythrocyte enzymes. In: Hereditary disorders of erythrocyte metabolism (BEUTLER, E., Ed.). City of Hope Symposium Series, Vol. 1, pp. 1—20. New York: Grune & Stratton 1967.

HERZENBERG, L. A., MINNA, J. D., HERZENBERG, L. A.: The chromosome region for immunoglobulin heavy-chains in the mouse. Cold Spring Harbor Symposia Quant. Biol. **32**, 181—186 (1967).

HOOD, L., GRAY, W. R., SANDERS, B. G., DREYER, W. J.: Light chain evolution. Cold Spring Harbor Symposia Quant. Biol. **32**, 133—146 (1967).

KOLER, R. D., BIGLEY, R. H., JONES, R. T., RIGAS, D. A., VANBELLINGHEN, P., THOMPSON, P.: Pyruvate kinase: Molecular differences between human red cell and leukocyte enzyme. Cold Spring Harbor Symposia Quant. Biol. **24**, 213—221 (1964).

MARKERT, C. L.: Cellular differentiation—An expression of differential gene function. In: Congenital malformations, pp. 163—174. New York: The International Medical Congress 1964.

NATVIG, J. B., KUNKEL, H. G., LITWIN, S. P.: Genetic markers of the heavy-chain subgroups of human gamma G globulin. Cold Spring Harbor Symposia Quant. Biol. **32**, 173—180 (1967).

PENHOET, E., RAJKUMAR, T., RUTTER, W. J.: Multiple forms of fructose diphosphate aldolase in mammalian tissues. Proc. Natl. Acad. Sci. US **56**, 1275—1282 (1966).

PLAGEMANN, P. G., GREGORY, K. F., WROBLEWSKI, F.: The electrophoretically distinct forms of mammalian lactic dehydrogenase. II. Properties and interrelationships of rabbit and human lactic dehydrogenase isozymes. J. Biol. Chem. **235**, 2282—2293 (1960).

POTTER, M., LIEBERMAN, R.: Genetic studies of immunoglobulins in mice. Cold Spring Harbor Symposia Quant. Biol. **32**, 203—209 (1967).

PUTNAM, F. W., TITANI, K., WIKLER, M., SHINODA, T.: Structure and evolution of kappa and lambda light chains. Cold Spring Harbor Symposia Quant. Biol. **32**, 9—29 (1967).

Chapter XIII

The Creation of a New Gene from a Redundant Duplicate of an Old Gene

The type of gene duplication which produced the group of isozyme genes contributed greatly to the evolution of increasingly complex organisms. These functionally diverged duplicated genes, however, still specify the same enzyme in that their products act upon the same substrate with the help of the same coenzyme. A, B and C-subunits of LDH of any vertebrate must still maintain either the identical active site sequence of 12 amino acids (— Val–Ile–Ser–Gly–Gly–Cys–Asn–Leu–Asp–Thr–Ala–Arg —), or a sequence very similar to the above, for this is the sequence which binds with NAD and recognizes pyruvate or lactate as the substrate (KAPLAN, 1965). Differences in the kinetic property of A, B and C-subunits must reflect amino acid substitutions which affect sites other than the active site of the polypeptide chain.